AD-A040 457 SOUTHERN METHODIST UNIV DALLAS TEX DEPT OF STATISTICS F/G 12/1
SOME CONSIDERATIONS IN THE EVALUATION OF ALTERNATE PREDICTION E--ETC(U)
1977 R F GUNST, R L MASON AF-AFOSR-2871-75
UNCLASSIFIED AFOSR-TR-77-0688 NL

| OF |
AD
A040457

END

DATE
FILMED
7-77

SOME CONSIDERATIONS IN THE EVALUATION OF ALTERNATE PREDICTION EQUATIONS

Richard F. Gunst
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

and

Robert L. Mason
Automotive Research Division
Southwest Research Institute
San Antonio, Texas 78284

Prediction equations constructed from multiple linear regression analyses are often intended for use in predicting response values throughout a region of the space of the predictor variables. Criteria for evaluating prediction equations, however, have generally concentrated attention on mean squared error properties of the estimated regression coefficients or on mean squared error properties of the predictor at the design points. If adequate prediction throughout a region of the space of predictor variables is the goal, neither of these criteria may be satisfactory in assessing the predictor. In this paper integrated mean squared error is used as a criterion to determine when the least squares, principal component, and ridge regression estimators of regression coefficients can produce satisfactory prediction equations in the presence of a multicollinear design matrix.

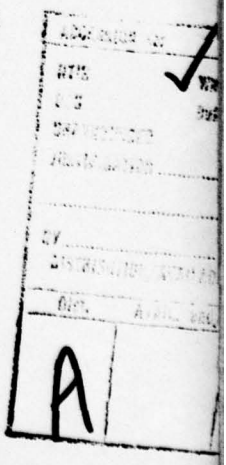
KEY WORDS

Integrated Mean Squared Error
Prediction Equations
Multicollinearity
Biased Estimation

1. INTRODUCTION

Box and Draper [1] espouse the use of an integrated mean squared error criterion to evaluate experimental designs proposed for use in fitting response surface models. Specifically, Box and Draper are concerned with the appropriate selection of design points $\underset{\sim}{X}_i' = (X_{i1}, X_{i2}, \ldots, X_{ip})$, $i = 1, 2, \ldots, n$, so that the mean squared error of the least squares prediction equation, integrated over an appropriate region of interest in the p design variables, is suitably small. Thus, experimental designs can be evaluated with respect to (i) the variance of the fitted model, (ii) bias incurred when an incorrect functional form is assumed between the response variable and the design variables, (iii) particular regions of interest of the design variables, and (iv) weighting functions that enable some regions of the predictor variables to influence the integrated mean squared error more heavily than others. The flexibility and intuitive appeal of integrated mean squared error has resulted in several subsequent papers evaluating both response surface designs (e.g. [2], [4], [5], [6], [7]) and estimators of the response function (e.g. [3], [10], [13], [14], [15], [16], [21]).

The purpose of this paper is to show that integrated mean squared error is a valuable tool in evaluating prediction equations arising from the use of different estimators of the unknown parameters in multiple linear regression models. The situation discussed in this paper differs from the one posed in most of the above articles in that we assume the data analyst has no control over the predictor (design) variables; i.e., the experimenter cannot select the values



of the predictor variables for which data on the response variable is to be obtained. This type of data is often characterized by the occurrence of multicollinearities among the predictor variables and consequent poor prediction when least squares prediction equations are employed (see, for example, [11], [19], and [20]). For these reasons, biased regression estimators have become very popular when regression data is multicollinear.

Hocking [11] and Gunst and Mason [8] reference much of the literature dealing with comparisons of estimators of regression coefficients. Overwhelmingly, these articles deal with comparisons using the mean squared errors of the estimated regression coefficients, although a few authors examine pointwise mean squared errors of the prediction equations at the design points. Yet the potential advantages of assessing a prediction equation using integrated mean squared error are many: the variances of the prediction equations are included in the assessment, biases due to the use of biased regression estimators and also due to misspecifications of the model can be evaluated, regions of the space of predictor variables for which one estimator has smaller integrated mean squared error than another can be identified, and unequal weightings can be assigned to regions of the space of predictor variables to reflect different requirements for accurate prediction.

Section 2 of this paper models the problem addressed in this paper and contrasts it with the one considered by Box and Draper. One important distinction noted between regression analysis and choosing an experimental design to estimate response surface models

is that, unlike the design problem, multicollinearities among the predictor variables in a regression analysis frequently cause the variance portion of the integrated mean squared error to be much larger than the bias due to using a biased estimator of the regression coefficients. In Sections 3, 4, and, 5, general expressions for integrated variance, integrated squared bias, and integrated mean squared error, respectively, are presented, along with specific results for models with two predictor variables. Section 6 briefly discusses estimation of integrated mean squared error and presents a numerical example. Conclusions and recommendations for further research are given in Section 7.

2. THE PROBLEM

We will concentrate attention in this article to multiple linear regression models of the following form:

$$\underline{Y} = \beta_0 \underline{1} + X \underline{\beta} + \underline{\varepsilon}, \tag{2.1}$$

where $\underline{Y}$ is an $(n \times 1)$ vector of response variables, β_0 is an unknown constant, $\underline{1}$ is an $(n \times 1)$ vector of ones, $X = [\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_p]$ is an $(n \times p)$ full column rank matrix of known nonstochastic predictor variables, $\underline{\beta}$ is a $(p \times 1)$ vector of unknown regression coefficients, and $\underline{\varepsilon}$ is an $(n \times 1)$ vector of random error terms with $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I)$. Except for the example discussed later in this section, we assume that model (2.1) has been correctly specified by the experimenter and that the columns of X have been standardized so that $\underline{X}_j' \underline{1} = 0$ and $\underline{X}_j' \underline{X}_j = 1$ for $j = 1, 2, \ldots, p$.

Consider a prediction equation of the form

$$\hat{Y}(\underline{u}) = \tilde{\beta}_0 + \underline{u}'\tilde{\underline{\beta}}, \tag{2.2}$$

where $\underline{u}' = (u_1, u_2, \ldots, u_p)$ is a vector of standardized (as in (2.1)) values of the p predictor variables at which a predicted value of the response variable is desired, and $\tilde{\beta}_0$ and $\tilde{\underline{\beta}}$ are estimators of the unknown constants in (2.1). If $\underline{X}_j' \underline{1} = 0$, i.e. the predictor variables are centered, then we will use $\tilde{\beta}_0 = \bar{Y}$.

If the prediction equation (2.2) is to be used for a range of values of the predictor variables, some measure of the adequacy of prediction throughout this region of the predictor variable space is needed to assess its efficacy. One such measure is integrated mean squared error, J, defined as

$$J = \int_R \cdots \int E\{\hat{Y}(\underline{u}) - E[Y(\underline{u})]\}^2 W(\underline{u}) d\underline{u}. \tag{2.3}$$

As defined in (2.3), integrated mean squared error incorporates the mean squared error of the prediction equation at the point $\underline{u}$, i.e. $E\{\hat{Y}(\underline{u}) - E[Y(\underline{u})]\}^2$, weighted by an appropriate function $W(\underline{u})$ and integrated over a region $R = R(\underline{u})$. This definition of integrated mean squared error can be adapted to discrete weight functions and models in which some predictor variables are functionally related to one another; however, we will restrict our attention to continuous predictor variables for simplicity (Helms [9] treats some of the complications of the more general definitions of J).

Box and Draper [1] analyzed in some detail the choice of an

experimental design for fitting a quadratic response surface

$$Y = \beta_0 + \beta_1 X_1 + \beta_{11} X_1^2 + \varepsilon \tag{2.4}$$

when it was incorrectly assumed that the response surface was linear, i.e.

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon. \tag{2.5}$$

We wish to discuss this example to point out the differences that occur when one can choose the design points and then estimate the regression coefficients versus the problems that arise when one cannot do so. Box and Draper assumed that the design points could be centered so that $\sum_{i=1}^{n} X_{i1} = 0$ and that the region of interest R could be chosen (through a scaling of X_1) to be $-1 \leq X_1 \leq 1$. They also chose the weight function to be constant throughout R; in particular, they let $W(\underline{u}) = n\sigma^{-2}(\int_{-1}^{1} du_1)^{-1} = n\sigma^{-2}/2$. Using least squares estimators of β_0 and β_1 it is then easily verified that (note we have not assumed that $\sum_{i=1}^{n} X_{i1}^2 = 1$ in this example)

$$J = V + B = (1 + 1/3c) + \alpha_{11}^2\{c^2 - 2c/3 + 1/5 + d^2/3c^2\}, \tag{2.6}$$

where the first term on the r.h.s. of (2.6) is the integrated variance (V) of the prediction equation, while the second term is the integrated squared bias (B), with $\alpha_{11}^2 = n\sigma^{-2}\beta_{11}^2$, $c = n^{-1}\sum_{i=1}^{n} X_{i1}^2$, and $d = n^{-1}\sum_{i=1}^{n} X_{i1}^3$.

In choosing the design points to minimize (2.6), Box and Draper noted that, regardless of the value of c, J would be smallest when

d = 0. In selecting the value of c to minimize J, however, one must specify a value for α_{11}^2. Alternatively, if V and B are restricted, unique values of c and α_{11}^2 can be found to minimize J. For example, if one considers a situation in which the variance and bias terms in (2.6) are equal, one can solve for the values of c and α_{11}^2 that minimize J subject to the restriction V = B.

Box and Draper investigated this further by solving for the minimizing values of $c^{\frac{1}{2}}$ for four cases: V = ∞, B = 0 ($c^{\frac{1}{2}} = \infty$); V = 4B ($c^{\frac{1}{2}} = 0.72$); V = B ($c^{\frac{1}{2}} = 0.62$); and V = 0, B = ∞ ($c^{\frac{1}{2}} = 0.58$). By noting the similarity of the values of $c^{\frac{1}{2}}$ in the last three cases versus $c^{\frac{1}{2}} = \infty$ for the first one, they reached the rather surprising conclusion that when the true model is quadratic but one assumes a linear one, designs that incorporate contributions from both variance and bias in the minimization of integrated mean squared error are very similar to those that ignore variance completely and minimize the integrated squared bias.

Through other examples Box and Draper and subsequent authors showed that this same conclusion (i.e. optimal designs that incorporate both integrated variance and integrated squared bias when minimizing J are close to the all bias designs) is true in a variety of response surface situations. The major distinction between these examples and a regression analysis is the inability of the experimenter in the latter instance to select the design points. In particular the data analyst performing a regression analysis generally cannot guarantee, as did Box and Draper, that the columns of X are mutually orthogonal or that odd sample moments of the X_j are zero.

The effects of the nonorthogonality of the columns of X on the conclusions of Box and Draper can be illustrated by a simple extension of the above example.

Suppose that, instead of considering the one variable model (2.4), the true model relates the response variable to two predictor variables as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \varepsilon. \tag{2.7}$$

Again, unknowingly, suppose the experimenter assumes the model is linear in the two predictor variables, i.e. he assumes

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon. \tag{2.8}$$

In order to make the comparison as similar as possible to the one variable example considered above, also assume that

$$\Sigma X_{i1} = \Sigma X_{i2} = \Sigma X_{i1}^2 X_{i2} = \Sigma X_{i1} X_{i2}^2 = \Sigma X_{i1}^3 = \Sigma X_{i2}^3 = 0$$

and (2.9)

$$n^{-1} \Sigma X_{i1}^2 = n^{-1} \Sigma X_{i2}^2 = c,$$

with all the summations taken from i = 1 to i = n. Finally, we do not assume $n^{-1} \Sigma X_{i1} X_{i2} = 0$, but that $n^{-1} \Sigma X_{i1} X_{i2} = c\, r_{12}$.

With these assumptions, and letting $R = \{(X_1, X_2) : -1 \leq X_i \leq 1, i = 1,2\}$ and $W(\underline{u}) = n\sigma^{-2}/4$,

$$J = V + B = (1 + 2/3c(1 - r_{12}^2)) + (\alpha_{11}^2 + \alpha_{22}^2)\{(c - 1/3)^2 + 4/45\} + 2\alpha_{11}\alpha_{22}(c - 1/3)^2,$$

where $\alpha_{jj} = n\sigma^{-2} \beta_{jj}$, j = 1,2. Further simplifying the problem by

letting $\alpha_{11} = \alpha_{22} = \alpha$ yields

$$J = V + B = (1 + 2/3c(1 - r_{12}^2)) + \alpha^2\{4(c - 1/3)^2 + 8/45\}. \quad (2.10)$$

If $r_{12} = 0$, (2.10) is the two variable analog of (2.6) (with d = 0).

The value of $c^{\frac{1}{2}}$ that minimizes (2.10) when $r_{12} = 0$ and V = B is 0.61 and $\alpha = 3.892$. With these values of $c^{\frac{1}{2}}$ and α, V = B = 2.788. As a function of r_{12}, the values of $c^{\frac{1}{2}}$ that minimize the integrated variance (2.10) when B = 2.788 are $c^{\frac{1}{2}} = 0.68(r_{12} = .90)$, $0.73(r_{12} = .95)$, $0.89(r_{12} = 0.99)$, and $1.23(r_{12} = .999)$. In addition, if $c^{\frac{1}{2}} = 0.61$ is used to construct an experimental design for this example but $r_{12} \neq 0$, the integrated squared bias, B, remains constant (since B is not a function of r_{12}) but the integrated variance becomes $V = 10.409(r_{12} = .90)$, $19.336(r_{12} = .95)$, $90.839(r_{12} = .99)$, and $895.342(r_{12} = .999)$. Finally, if B = 2.788, the values of $c^{\frac{1}{2}}$ needed to insure that V = B are $c^{\frac{1}{2}} = 0.61(r_{12} = 0)$, $1.40(r_{12} = .90)$, $1.96(r_{12} = .95)$, $4.33(r_{12} = .99)$, and $13.66(r_{12} = .999)$.

Thus in selecting an experimental design for fitting (2.8) when (2.7) is the correct model, the integrated variance cannot be ignored if r_{12} is close to 1.0. Values of r_{12} near 1.0 frequently arise with regression data and can result in an extremely large integrated variance for the least squares prediction equation, even if the model is correctly specified. Thus, although biased regression estimators contribute nonzero integrated squared biases to J, the reduction in integrated variance over the least squares estimator can result in an overall reduction in integrated mean squared error with multicollinear

data. The magnitude of a reduction in integrated mean squared error over least squares, if any, will depend on the region of interest and the weight function used in (2.3), as well as on the particular biased estimator employed.

3. INTEGRATED VARIANCE

The integrated mean squared error (2.3) of a prediction equation can be partitioned into two components: integrated variance, V, and integrated squared bias, B, where

$$V = \int_R \cdots \int \mathrm{Var}\{\hat{Y}(\underline{u})\}\, W(\underline{u})\, d\underline{u} \tag{3.1}$$

and

$$B = \int_R \cdots \int \{E[\hat{Y}(\underline{u})] - E[Y(\underline{u})]\}^2\, W(\underline{u})\, d\underline{u}. \tag{3.2}$$

The three regression estimators to be compared in this article are the ordinary least squares (LS), simple ridge regression (RR), and principal component (PC) estimators defined in equations (3.4) - (3.6). To facilitate the evaluation of the associated prediction equations, define the latent roots of $X'X$ by $\ell_1 \leq \ell_2 \leq \cdots \leq \ell_p$ and the corresponding orthonormal latent vectors by $\underline{V}_1, \underline{V}_2, \ldots, \underline{V}_p$. It is well-known, then, that

$$X'X = \sum_{j=1}^{p} \ell_j \underline{V}_j \underline{V}_j' \quad \text{and} \quad (X'X)^{-1} = \sum_{j=1}^{p} \ell_j^{-1} \underline{V}_j \underline{V}_j'. \tag{3.3}$$

Using (3.3), the LS estimator of $\underline{\beta}$ can be written as

$$\hat{\underline{\beta}}_{LS} = (X'X)^{-1} X'\underline{Y} = \sum_{j=1}^{p} \ell_j^{-1} C_j \underline{V}_j, \tag{3.4}$$

where $C_j = \underline{V}_j'X'\underline{Y}$. The PC estimator we will examine deletes terms from (3.4) corresponding to multicollinearities among the predictor variables as indicated by the presence of small latent roots of X'X (see Marquardt [18], Mansfield [17], and Gunst and Mason [8] for a more complete discussion of this estimator, including justification for deleting terms solely on the basis of the magnitudes of the latent roots of X'X). If the terms corresponding to r small latent roots are deleted from (3.4) the resulting PC estimator of $\underline{\beta}$ is

$$\hat{\underline{\beta}}_{PC} = \sum_{j=r+1}^{p} \ell_j^{-1} C_j \underline{V}_j. \tag{3.5}$$

Finally, for $k > 0$ the RR estimator (Hoerl and Kennard [12]) is given by

$$\hat{\underline{\beta}}_{RR} = (X'X + kI)^{-1} X'\underline{Y} = \sum_{j=1}^{p} (\ell_j + k)^{-1} C_j \underline{V}_j. \tag{3.6}$$

In deriving the integrated mean squared error of the RR estimator we assume that k is a fixed constant, i.e. nonstochastic. Although in practice k is typically selected according to a stochastic method (e.g. Ridge Trace, k estimated using $\hat{\underline{\beta}}_{LS}$ and $\hat{\sigma}^2$, etc.), nonrandom selection rules utilizing only characteristics of X are conceivable. One such procedure is introduced in Section 6.

Assuming that (2.1) is the correct model, the integrated variances of the above three estimators are, respectively,

$$V_{LS} = n^{-1}\sigma^2 + \sigma^2 \operatorname{tr}\{(X'X)^{-1}\ddagger\} = n^{-1}\sigma^2 + \sigma^2 \sum_{j=1}^{p} \ell_j^{-1} \underline{V}_j' \ddagger \underline{V}_j \tag{3.7}$$

$$V_{PC} = n^{-1}\sigma^2 + \sigma^2 \operatorname{tr}\{(X'X)^{-}\ddagger\} = n^{-1}\sigma^2 + \sigma^2 \sum_{j=r+1}^{p} \ell_j^{-1} \underline{V}_j' \ddagger \underline{V}_j \tag{3.8}$$

and

$$V_{RR} = n^{-1}\sigma^2 + \sigma^2 \operatorname{tr}\{(X'X + kI)^{-1} X'X(X'X + kI)^{-1}\nabla\}$$

$$= n^{-1}\sigma^2 + \sigma^2 \sum_{j=1}^{p} \ell_j(\ell_j + k)^{-2} \underline{V}_j' \nabla \underline{V}_j, \qquad (3.9)$$

where $(X'X)^{-} = \sum_{j=r+1}^{p} \ell_j^{-1} \underline{V}_j \underline{V}_j'$ and ∇ is the second order moment matrix of the weight function in the region R:

$$\nabla = \int_R \ldots \int \underline{u}\underline{u}' \, W(\underline{u}) d\underline{u}. \qquad (3.10)$$

Note immediately that $V_{PC} < V_{LS}$ for $r \geq 1$ and $V_{RR} < V_{LS}$ for $k > 0$; i.e. both PC and RR result in reductions in integrated variance over LS. With multicollinear data these reductions can be quite large since the prediction equations using the biased estimators either eliminate (PC) or dampen (RR) the first few terms of (3.7) that contain the largest values of ℓ_j^{-1}. The magnitude of V_{PC} relative to V_{RR} depends on the number of terms, r, deleted by the PC estimator of $\underline{\beta}$ and the value of k selected for the RR estimator.

One of the simplest forms of ∇ in (3.10) occurs when R is a symmetric region about $\underline{u} = \underline{0}$ in each u_j and the following two conditions hold for the weight function $W(\underline{u})$:

(i) $W(\underline{u})$ is normalized so that $\int_R \ldots \int W(\underline{u}) d\underline{u} = 1$ and is an even function of each u_j; and

(ii) $\int_R \ldots \int u_j^2 \, W(\underline{u}) d\underline{u} = \tau$, a constant, for $j = 1,2, \ldots, p$.

For example, many uniform, triangular, and exponential weight functions satisfy these requirements. When these conditions are valid

$$\mathfrak{X} = \tau I, \tag{3.11}$$

and the integrated variances become, respectively,

$$V_{LS} = n^{-1}\sigma^2 + \sigma^2\tau \sum_{j=1}^{p} \ell_j^{-1} \tag{3.12}$$

$$V_{PC} = n^{-1}\sigma^2 + \sigma^2\tau \sum_{j=r+1}^{p} \ell_j^{-1} \tag{3.13}$$

and

$$V_{RR} = n^{-1}\sigma^2 + \sigma^2\tau \sum_{j=1}^{p} \ell_j(\ell_j + k)^{-2}. \tag{3.14}$$

Now to further illuminate the tradeoffs in integrated variance among these three prediction equations, we consider the case of $p = 2$ predictor variables and $\mathfrak{X}$ defined as in (3.11). If r_{12} again denotes the "correlation" between the n observations on the two predictor variables, then $\ell_1 = 1 - r_{12}$ and $\ell_2 = 1 + r_{12}$ (assuming w.l.o.g. that $r_{12} > 0$). If the PC estimator deletes the term corresponding to the smallest latent root, ℓ_1, then

$$V^*_{LS} = 2(1 - r_{12}^2)^{-1}, \tag{3.15}$$

$$V^*_{PC} = (1 + r_{12})^{-1}, \tag{3.16}$$

and

$$V^*_{RR} = (1 - r_{12})(1 - r_{12} + k)^{-2} + (1 + r_{12})(1 + r_{12} + k)^{-2}, \tag{3.17}$$

where, for each predictor, $V^* = (V - n^{-1}\sigma^2)/\sigma^2\tau$. Figure 1 contains graphs of V^*_{LS}, V^*_{PC}, and V^*_{RR} (for several values of k) as a function

of r_{12}.

[Insert Figure 1]

The comparisons among the predictors that are evident from Figure 1 include the following:

(i) LS has uniformly larger integrated variance than PC and RR, with V^*_{LS} asymptotically unbounded as $r_{12} \to 1$;

(ii) for k approximately 0.4 or less the integrated variance of RR is larger than that of PC except for values of r_{12} extremely close to 1.0.

Thus for p = 2 predictor variables and $\Sigma = \tau I$, both biased estimators greatly reduce the integrated variance over that obtainable by LS except for RR when k is small and, simultaneously, r_{12} is not close to 1.0; i.e. except when k is small and the two predictor variables are not severly multicollinear. Comparing PC with RR reveals that V^*_{PC} is generally smaller than V^*_{RR} unless k is relatively large or r_{12} is extremely close to 1.0.

Another comparison between V_{PC} and V_{RR} for p = 2 is presented in Figure 2, which displays the regions for which $V_{PC} \leq V_{RR}$ and $V_{PC} > V_{RR}$ as a function of r_{12}. Again this graph shows that for small k $V_{PC} \leq V_{RR}$ unless r_{12} is large and that $V_{PC} > V_{RR}$ for large k over a wide range of r_{12}.

4. INTEGRATED SQUARED BIAS

The integrated squared biases of the prediction equation (2.2) using coefficient estimators (3.4) - (3.6) are, respectively,

$$B_{LS} = 0, \tag{4.1}$$

$$B_{PC} = \mathrm{tr}\{V_r V_r' \underline{\beta}\underline{\beta}' V_r V_r' \Psi\}$$

$$= \underline{\beta}' V_r V_r' \Psi V_r V_r' \underline{\beta}, \tag{4.2}$$

and

$$B_{RR} = k^2 \mathrm{tr}\{V(L + kI)^{-1} V' \underline{\beta}\underline{\beta}' V(L + kI)^{-1} V' \Psi\}$$

$$= k^2 \underline{\beta}' V(L + kI)^{-1} V' \Psi V(L + kI)^{-1} V' \underline{\beta}, \tag{4.3}$$

where $V_r = [\underline{V}_1, \underline{V}_2, \ldots, \underline{V}_r]$ and $L = \mathrm{diag}(\ell_1, \ell_2, \ldots, \ell_p)$. Although the LS estimator contributes no bias to the integrated mean squared error, the bias contribution of PC and RR to their respective integrated mean squared errors can be small enough to net great reductions over J_{LS} due to the large reductions in integrated variance. This is especially true for severely multicollinear data and weight functions $W(\underline{u})$ that give smallest weights to regions that have the same multicollinearities as those in the matrix of predictor variables, X.

If we again examine the characteristics of regions and weight functions yielding $\Psi = \tau I$, (4.2) and (4.3) reduce to

$$B_{PC} = \tau \sum_{j=1}^{r} (V_j' \underline{\beta})^2 \tag{4.4}$$

and

$$B_{RR} = \tau k^2 \sum_{j=1}^{p} (\ell_j + k)^{-2} (\underline{V}_j' \underline{\beta})^2. \tag{4.5}$$

It is readily apparent from (4.4) and (4.5) that $B_{PC} \leq B_{RR}$ when $D_p \leq 0$ and $B_{PC} > B_{RR}$ when $D_p > 0$, where

$$D_p = \sum_{j=1}^{r} (1 - a_j^2) Z_j^2 - \sum_{j=r+1}^{p} a_j^2 Z_j^2, \tag{4.6}$$

$Z_j = \underset{\sim}{V}_j'\underset{\sim}{\beta}$, and $a_j = k/(\ell_j + k)$. The results of the previous section suggest that large values of k yield an integrated variance for ridge regression that is smaller than the integrated variance for principal components. But there is a tradeoff in integrated squared bias since large values of k imply that $a_j \approx 1$ for latent roots defining multicollinearities (i.e. for j = 1, 2, ..., r) and hence that

$$D_p \approx - \sum_{j=r+1}^{p} a_j^2 Z_j^2 < 0.$$

This property is of course weakened if some of the Z_j, j = 1, 2, ..., r, are large enough to offset the closeness of the corresponding a_j to 1.

For p = 2 predictor variables and $\underset{\sim}{\Sigma} = \tau I$,

$$B_{PC} = \tau(\beta_1 - \beta_2)^2/2 \tag{4.7}$$

and

$$B_{RR} = \tau k^2\{(\beta_1-\beta_2)^2(1-r_{12}+k)^{-2} + (\beta_1+\beta_2)^2(1+r_{12}+k)^{-2}\}/2, \tag{4.8}$$

again assuming r = 1 and $r_{12} > 0$. As a function of a_1, a_2, Z_1, and Z_2, these expressions are

$$B_{PC} = \tau Z_1^2 \text{ and } B_{RR} = \tau\{a_1^2 Z_1^2 + a_2^2 Z_2^2\}. \tag{4.9}$$

Figure 3 depicts regions of the (Z_1, Z_2) - plane for which $B_{PC} < B_{RR}$ for p = 2 predictor variables, $\Sigma = \tau I$, $r_{12} = 0.95$, and four choices of k. Similar regions occur for other values of r_{12}, only the slopes of the lines change. The locus of points for which $B_{PC} = B_{RR}$ is two lines passing through the origin with slopes $\pm a_2^{-1}(1 - a_1^2)^{\frac{1}{2}}$.

[Insert Figure 3]

Examination of a prediction equation with p = 3 predictor variables reveals some general characteristics of integrated bias comparisons between PC and RR. In this case (recall equation (4.6))

$$D_3 = \begin{cases} (1 - a_1^2)Z_1^2 - a_2^2Z_2^2 - a_3^2Z_3^2 & r = 1 \\ \\ (1 - a_1^2)Z_1^2 + (1 - a_2^2)Z_2^2 - a_3^2Z_3^2 & r = 2 \end{cases} \tag{4.10}$$

The general shape of the region defined by (4.10) when PC deletes r = 1 terms from (3.4) to obtain (3.5) is that of an elliptical cone centered on the Z_1 axis. Outside this cone $B_{PC} < B_{RR}$, while inside it $B_{PC} > B_{RR}$; i.e., $B_{PC} < B_{RR}$ unless Z_1^2 is sufficiently large. This comparison generalizes to an arbitrary number of predictor variables, p, for r = 1.

When p = 3 and r = 2, D_3 indicates that the general shape of the region comparing B_{PC} and B_{RR} is again characterized by an elliptical cone, now centered on the Z_3 axis. Inside this cone $B_{PC} < B_{RR}$, while outside it $B_{PC} > B_{RR}$. So $B_{PC} < B_{RR}$ for arbitrary Z_3^2 if neither Z_1^2 nor Z_2^2 is too large. This conclusion remains valid for arbitrary p provided r = p - 1.

5. INTEGRATED MEAN SQUARED ERROR

The tradeoffs in integrated variance and integrated squared bias that were uncovered in the previous two sections can be evaluated by considering the integrated mean squared errors of the prediction equations:

$$J_{LS} = n^{-1}\sigma^2 + \sigma^2 \operatorname{tr}\{(X'X)^{-1}\Psi\}, \tag{5.1}$$

$$J_{PC} = n^{-1}\sigma^2 + \sigma^2 \operatorname{tr}\{(X'X)^{-}\Psi\} + \underline{\beta}'V_r V_r'\Psi V_r V_r'\underline{\beta}, \tag{5.2}$$

$$J_{RR} = n^{-1}\sigma^2 + \sigma^2 \operatorname{tr}\{(X'X + kI)^{-1}X'X(X'X + kI)^{-1}\Psi\}$$

$$+ k^2\underline{\beta}'V(L + kI)^{-1}V'\Psi V(L + kI)^{-1}V'\underline{\beta}. \tag{5.3}$$

Rather than attempting a complicated comparison of expressions (5.1) - (5.3), we again simplify the problem by specifying that $\Psi = \tau I$. Then,

$$J_{LS} = \sigma^2\{n^{-1} + \tau \sum_{j=1}^{p} \ell_j^{-1}\}, \tag{5.4}$$

$$J_{PC} = \sigma^2\{n^{-1} + \tau \sum_{j=r+1}^{p} \ell_j^{-1}\} + \tau \sum_{j=1}^{r} z_j^2, \tag{5.5}$$

and $$J_{RR} = \sigma^2\{n^{-1} + \tau \sum_{j=1}^{p} \ell_j(\ell_j + k)^{-2}\} + \tau k^2 \sum_{j=1}^{p} (\ell_j + k)^{-2} z_j^2. \tag{5.6}$$

Examination of (5.4) and (5.5) reveals that $J_{LS} \leq J_{PC}$ when $E_p \leq 0$ and $J_{LS} > J_{PC}$ when $E_p > 0$, where

$$E_p = \sum_{j=1}^{r} (\ell_j^{-1}\sigma^2 - z_j^2). \tag{5.7}$$

Comparing (5.4) with (5.6) reveals that $J_{LS} \leq J_{RR}$ when $F_p \leq 0$ and $J_{LS} > J_{RR}$ when $F_p > 0$, where

$$F_p = \sum_{j=1}^{p} (b_j - a_j^2 Z_j^2), \tag{5.8}$$

$$b_j = k\sigma^2(2\ell_j + k)/[\ell_j(\ell_j + k)^2],$$

and $a_j = k/(\ell_j + k)$ as in Section 4. Finally, from (5.5) and (5.6), $J_{PC} \leq J_{RR}$ when $D_p \leq d_p$ and $J_{PC} > J_{RR}$ when $D_p > d_p$, where D_p is defined in (4.6) and

$$d_p = (V_{RR} - V_{PC})/\tau$$

$$= \sum_{j=r+1}^{p} b_j - \sigma^2 \sum_{j=1}^{r} \ell_j(\ell_j + k)^{-2}. \tag{5.9}$$

For the two variable prediction equation considered in the previous sections,

$$J^*_{LS} = 2\sigma^2(1 - r_{12}^2)^{-1}, \tag{5.10}$$

$$J^*_{PC} = \sigma^2(1 + r_{12})^{-1} + Z_1^2, \tag{5.11}$$

and $$J^*_{RR} = \sigma^2\{(1-r_{12})(1-r_{12}+k)^{-2} + (1+r_{12})(1+r_{12}+k)^{-2}\}$$

$$+ k^2\{(1-r_{12}+k)^{-2}Z_1^2 + (1+r_{12}+k)^{-2}Z_2^2\}, \tag{5.12}$$

where $J^* = (J - n^{-1}\sigma^2)/\tau$. Figures 4 and 5 pictorially reveal the combined effects of integrated variance and integrated squared bias

of the two variable predictors by plotting (5.10) - (5.12) as a function of Z_1^2, for $\sigma^2 = 1$ (hence Z_j^2 measures the magnitude of $(\underset{-}{V}_j'\underset{-}{\beta})^2$ relative to σ^2). Two values of r_{12} are used to indicate the changes in the curves as r_{12} is changed, and only the curves for $Z_2^2 = 0$ are depicted. Nonzero values of Z_2^2 increase the intercept values for J^*_{RR} but leave the curves for J^*_{LS} and J^*_{PC} unchanged. So this is a "worst case" comparison of J^*_{LS} and J^*_{PC} with J^*_{RR}.

[Insert Figures 4 and 5]

In general, these figures support the contention that reductions in integrated mean squared error over LS are possible with either biased estimator provided that Z_1^2 is not too large relative to σ^2 (and to a lesser extent, provided that Z_2^2 is not too large for RR). Due to the magnitude of V_{LS} for the stronger multicollinearity, $r_{12} = 0.99$, substantial reductions in integrated mean squared error are seen to be possible with the biased estimators when the predictor variables are extremely multicollinear. The comparison of J^*_{PC} and J^*_{RR} indicates that $J^*_{PC} \leq J^*_{RR}$ for smaller values of Z_1^2, particularly for small values of k. Large values of Z_1^2 or large selections of k result in smaller integrated mean squared error for RR than PC, provided a large value of Z_2^2 doesn't compensate for these reductions.

6. ESTIMATION

Helms [10] and Park [22] employ integrated mean squared error criteria to assess different least squares prediction equations that arise due to attempts to select acceptable subsets of the original p

predictor variables for use in a final model. Helms [10] uses known characteristics of the predictor variables to define ξ and then estimates the integrated variances of the subset models as a (biased) mimic of the corresponding integrated mean squared errors. Park [22], analyzing a p = 3 variable model, determines ξ by defining $W(\underline{u})$ to be a uniform weight function and R to be the unit cube. He then estimates integrated mean squared error by evaluating J for various subsets of the full set of predictor variables using the least squares estimates of the parameters from the full model. Both of these procedures for estimating integrated mean squared error yield biased estimators of J.

In this section we will develop an alternate approach for estimating integrated mean squared error. We will use characteristics of the data and each of the three regression estimators discussed in this paper to define a region of prediction R and an estimator $\hat{J}$ so that $\hat{J}$ is an unbiased estimator of the corresponding J. An example illustrating some of the characteristics of this estimation scheme concludes the section.

Since we are primarily concerned with multicollinear data, a transformation to the principal axes of X'X rather than using the original coordinate system allows R to be defined to reflect anomalies in the data. For example, if X is severely multicollinear there is very little information in the p dimensional space of the predictor variables in directions defined by the latent vectors corresponding to the small latent roots of X'X. The region of prediction, R, should be chosen to reflect such characteristics

in the data; otherwise one is in danger of extrapolation with the predictor. So let

$$\underline{t} = V'\underline{u}, \tag{6.1}$$

where $\underline{u}$ is again a vector of standardized values of the p predictor variables in the original coordinate system, and $\underline{t}$ represents this same point in the orthogonal coordinate system defined by the p latent vectors of X'X; i.e., $t_j = \underline{V}_j'\underline{u}$ $j = 1, 2, \ldots, p$. Let $W^*(\underline{t})$ and R^* denote weight functions and regions of the predictor variables, respectively, in the transformed coordinate system.

Consider now the use of a rectangular region of interest in the transformed space which is defined by

$$R^* = \{\underline{t} : -s_j \leq t_j \leq s_j,\ j = 1, 2, \ldots, p\}, \tag{6.2}$$

where $s_j \geq 0$ will be defined for each estimator and represents the limits imposed on the use of a prediction equation in each direction of the transformed space of predictor variables. For illustrative purposes, a uniform weight function, $W^*(\underline{t})$, will be used to discuss the estimation of J although, as mentioned in Section 3, many other weight functions behave similarly. Accordingly, define

$$W^*(\underline{t}) = \begin{cases} \prod_{j=1}^{p} (2s_j)^{-1} & \underline{t} \in R^* \\ 0 & \underline{t} \notin R^*. \end{cases} \tag{6.3}$$

From (6.2) and (6.3),

$$\underset{\sim}{\Phi} = \int_{R^*} \cdots \int \underset{\sim}{t}\underset{\sim}{t}'W^*(\underset{\sim}{t})d\underset{\sim}{t} = \text{diag}(s_1, s_2, \ldots, s_p). \tag{6.4}$$

Noting that in the transformed space $(X'X)^{-1} = \text{diag}(\ell_1^{-1}, \ell_2^{-1}, \ldots, \ell_p^{-1})$, we find from (5.1) that

$$J_{LS} = n^{-1}\sigma^2 + \sigma^2 \sum_{j=1}^{p} \ell_j^{-1} s_j, \tag{6.5}$$

and, hence, that an unbiased estimator of (6.5), regardless of the choice of the s_j (provided they are not random variables), is

$$\hat{J}_{LS} = n^{-1}\hat{\sigma}^2 + \hat{\sigma}^2 \sum_{j=1}^{p} \ell_j^{-1} s_j, \tag{6.6}$$

where $\hat{\sigma}^2 = \text{MSE}$ is the usual unbiased estimator of σ^2.

Similarly, in the transformed space $(X'X)^{-} = \text{diag}(0, 0, \ldots, 0, \ell_{r+1}^{-1}, \ldots, \ell_p^{-1})$ and (5.2) becomes

$$J_{PC} = n^{-1}\sigma^2 + \sigma^2 \sum_{j=r+1}^{p} \ell_j^{-1} s_j + \sum_{j=1}^{r} (\underset{\sim}{V}_j'\underset{\sim}{\beta})^2 s_j. \tag{6.7}$$

We again estimate $n^{-1}\sigma^2$ by $n^{-1}\hat{\sigma}^2$, but rather than use $\hat{\sigma}^2$ in the second term of (6.7), we will use MSE_{PC}, defined as

$$\begin{aligned} \text{MSE}_{PC} &= \{\underset{\sim}{Y}'(I - n^{-1}\underset{\sim}{1}\underset{\sim}{1}' - X(X'X)^{-}X')\underset{\sim}{Y}\}/(n-p-1+r) \\ &= \{(n-p-1)\hat{\sigma}^2 + \sum_{j=1}^{r} \ell_j^{-1} C_j^2\}/(n-p-1+r), \end{aligned}$$

i.e. let

$$\hat{J}_{PC} = n^{-1}\hat{\sigma}^2 + \text{MSE}_{PC} \sum_{j=r+1}^{p} \ell_j^{-1} s_j. \tag{6.8}$$

The bias of $\hat{J}_{PC}$ is

$$E[\hat{J}_{PC}] - J_{PC} = \{\sum_{j=r+1}^{p} \ell_j^{-1} s_j\}\{\sum_{j=1}^{r} \ell_j (\underset{-}{V}_j'\underset{-}{\beta})^2/(n-p-1+r)\}$$

$$- \sum_{j=1}^{r} (\underset{-}{V}_j'\underset{-}{\beta})^2 s_j .$$

If we now restrict the region of prediction R* so that

$$s_j = \begin{cases} (n-p-1+r)^{-1}\ell_j & j = 1, 2, \ldots, r \\ (p-r)^{-1}\ell_j & j = r+1, r+2, \ldots, p \end{cases} \tag{6.9}$$

then $E[\hat{J}_{PC}] = J_{PC}$. Note that (6.9) does restrict prediction most in directions for which there is little information on the predictor variables (i.e. directions defined by latent vectors corresponding to small latent roots) and least in directions for which there is the most information.

The integrated mean squared error of the RR estimator using (6.4) is given by

$$J_{RR} = n^{-1}\sigma^2 + \sigma^2 \sum_{j=1}^{p} \ell_j(\ell_j+k)^{-2} s_j + k^2 \sum_{j=1}^{p} (\ell_j+k)^{-2} (\underset{-}{V}_j'\underset{-}{\beta})^2 s_j, \tag{6.10}$$

for nonstochastic choices of k. Define (note that this is not the usual mean squared error definition of the ridge estimator)

$$MSE_{RR} = \{\underset{-}{Y}'(I-n^{-1}\underset{-}{1}\underset{-}{1}'-X(X'X+kI)^{-1}\underset{-}{X}')\underset{-}{Y}\}/(n-1-\sum_{j=1}^{p} \ell_j(\ell_j+k)^{-1})$$

$$= \{(n-p-1)\hat{\sigma}^2 + k\sum_{j=1}^{p} \ell_j^{-1}(\ell_j+k)^{-1} C_j^2\}/(n-1-\sum_{j=1}^{p} \ell_j(\ell_j+k)^{-1}).$$

We estimate J_{RR} with

$$\hat{J}_{RR} = n^{-1}\hat{\sigma}^2 + MSE_{RR}\sum_{j=1}^{p} \ell_j(\ell_j+k)^{-2}s_j. \tag{6.11}$$

This is an unbiased estimator of J_{RR} if we restrict R* so that

$$s_j = k^{-1}\ell_j(\ell_j + k)/(n - 1 - \sum_{j=1}^{p} \ell_j(\ell_j + k)^{-1}), \tag{6.12}$$

and also use

$$k = p/(n - 1). \tag{6.13}$$

It is especially important to observe that in obtaining an unbiased estimator of J_{RR}, a nonstochastic rule for selecting the ridge shrinkage parameter k resulted.

To illustrate the use of these estimators we will examine the nine-variable data analyzed by Webster, Gunst, and Mason [23]. A least squares backward elimination of this data resulted in a final predictor involving four of the nine predictor variables: X_1, X_4, X_6, and X_8, with X_1 and X_4 having a large pairwise multicollinearity ($r_{14} = 0.978$). Using a latent root regression backward elimination procedure (Webster, Gunst, and Mason [23, 24]), the final predictor contains only two predictor variables, X_6 and X_9, which do not appear strongly multicollinear ($r_{69} = 0.143$). Both these subset predictors appear to be reasonably adequate predictors of the n = 15 observed responses (coefficients of determination for the two models are 0.80 and 0.75, respectively, and the residual mean squared errors are 2.40 and 2.49, while the corresponding statistics for the full model are 0.83 and 4.12).

Tables 1 and 2 contain the statistics used in the evaluation of the integrated mean squared errors of the prediction equations for each of the above subset models. The s_j values for RR and PC were obtained from (6.12) and (6.9), respectively, while those for LS were chosen to be the smaller (in absolute value) of the upper and lower bounds on the observed $t_j = \underset{-j}{V}'\underset{-i}{u}$, i = 1, 2, ..., n (observe that these values are merely the observed principal components of X). This choice of s_j for LS insures that we do not attempt to extrapolate along the axes in the transformed space.

TABLE 1. STATISTICS FOR THE TWO VARIABLE MODEL: X_6, X_9.

		Values of s_j			Range on t_j	
j	ℓ_j	RR	PC	LS	LOWER	UPPER
1	.8567	.4892	.0659	.3521	-.4240	.3521
2	1.1433	.8399	1.1433	.3927	-.3927	.6324

$\hat{J}_{LS}$ = 2.0406 $\hat{J}_{PC}$ = 7.8419 $\hat{J}_{RR}$ = 3.5931 (k=0.1429)

TABLE 2. STATISTICS FOR THE FOUR VARIABLE MODEL: X_1, X_4, X_6, X_8.

		Values of s_j			Range on t_j	
j	ℓ_j	RR	PC	LS	LOWER	UPPER
1	.0115	.0010	.0010	.0439	-.0708	.0439
2	.2355	.0362	.0785	.1718	-.2771	.1718
3	.8574	.2895	.2858	.2715	-.4816	.2715
4	2.8956	2.7210	.9653	.8030	-.8030	.8454

$\hat{J}_{LS}$ = 12.5227 $\hat{J}_{PC}$ = 3.1877 $\hat{J}_{RR}$ = 4.1786 (k=0.2857)

Consider first the two variable model. The region of prediction for PC is greatly distorted when comparison is made with the range on t_j for each of the two dimensions. The region of prediction for RR allows a great deal of extrapolation in the second dimension. The estimates of J indicate that the LS prediction equation should be preferred over PC and RR, with the PC predictor clearly inferior to the other two. These results are not especially surprising since there is no strong multicollinearity in this model.

The statistics presented in Table 2 point out the advantages of using either biased estimator of $\underline{\beta}$ in a prediction equation when the data is multicollinear. The regions of prediction for PC and RR are conservative in the first two dimensions when compared with the range on the t_j. The range of prediction for RR in the last dimension is quite anticonservative. The estimated integrated mean squared errors of PC and RR indicate that if prediction is confined to the regions in Table 2, both biased predictors greatly reduce the overall variability of the LS predictor.

7. SUMMARY AND RECOMMENDATIONS

The example in the previous section suggests several questions which need to be answered before a comparison of LS, PC, and RR prediction equations can be regarded as conclusive. First, the restriction of an unbiased estimator of J may be detrimental to the evaluation of the ridge prediction equation. The values of s_j for the largest dimension in each subset model was much larger than the bounds of the data in these directions. Yet a smaller

value of s_j renders $\hat{J}_{RR}$ biased for J_{RR}. Further investigations on the desirability of unbiased estimators of J are needed to resolve this problem.

In addition to the estimation problems, additional choices of weight functions should be considered. For example, letting $\mathbf{\Sigma} = X'X$ deserves consideration, as Helms [9] argues. Finally, how should comparisons be made if the model is misspecified? These problems are currently under investigation and additional results will be reported in the near future.

The unresolved questions just raised do not detract from the main thrust of this paper: integrated mean squared error is a flexible tool for evaluating competing prediction equations. The general formulation of applying this criterion to least squares, principal component, and ridge regression prediction equations has been presented for a correctly specified model, and comparisons have been made among the predictions for a general class of weight functions over a specific region of interest. By allowing the regions of prediction to vary, unbiased estimators of the integrated mean squared error of the three prediction equations were obtained and a numerical example illustrating the use of the procedures was discussed.

8. ACKNOWLEDGMENTS

This research was supported in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-75-2871.

Figure 1.
Integrated Variances as a Function of r_{12}

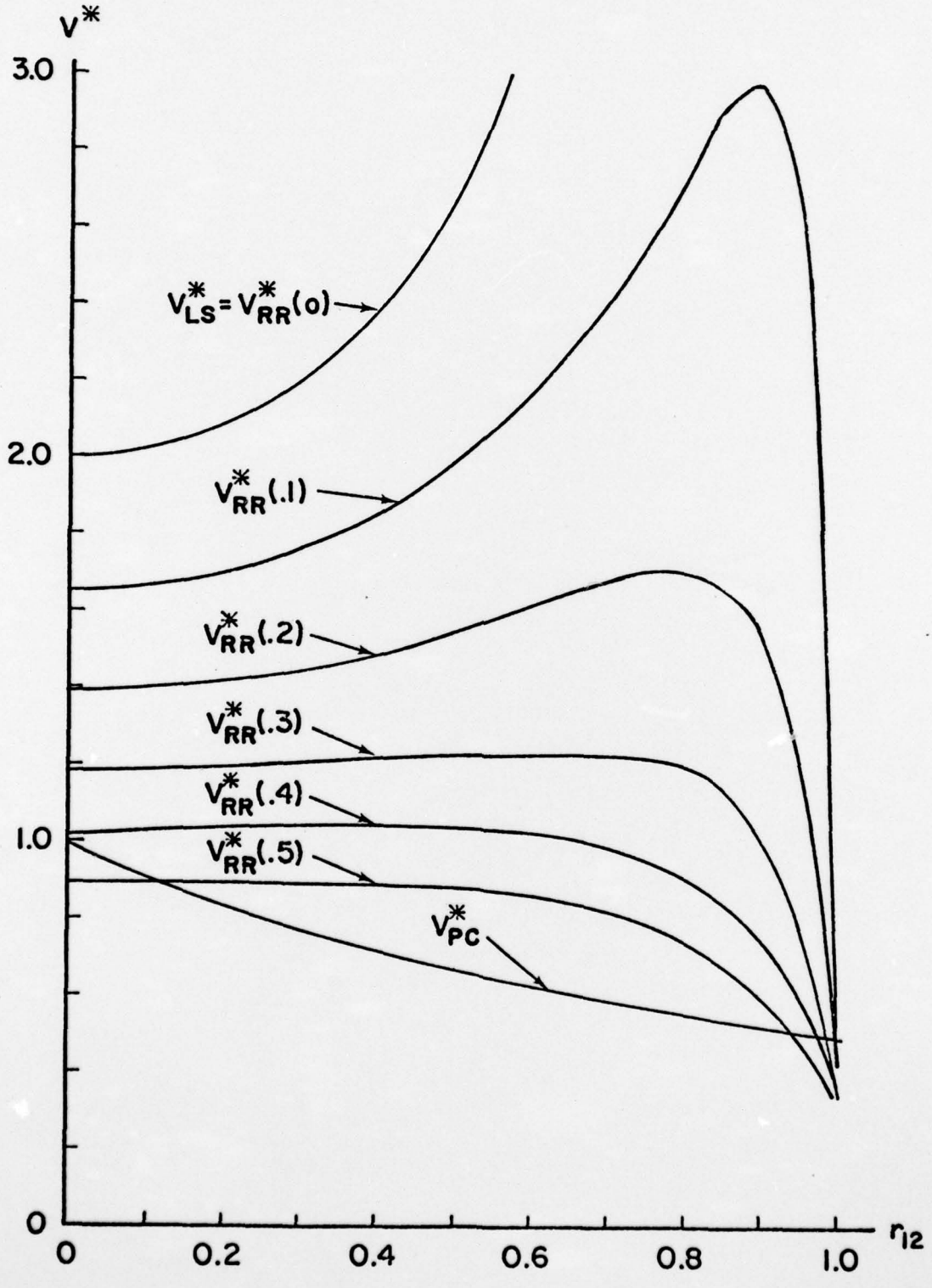


Figure 2.
Comparison of Integrated Variance of PCR and RR
for Two Regressor Variables

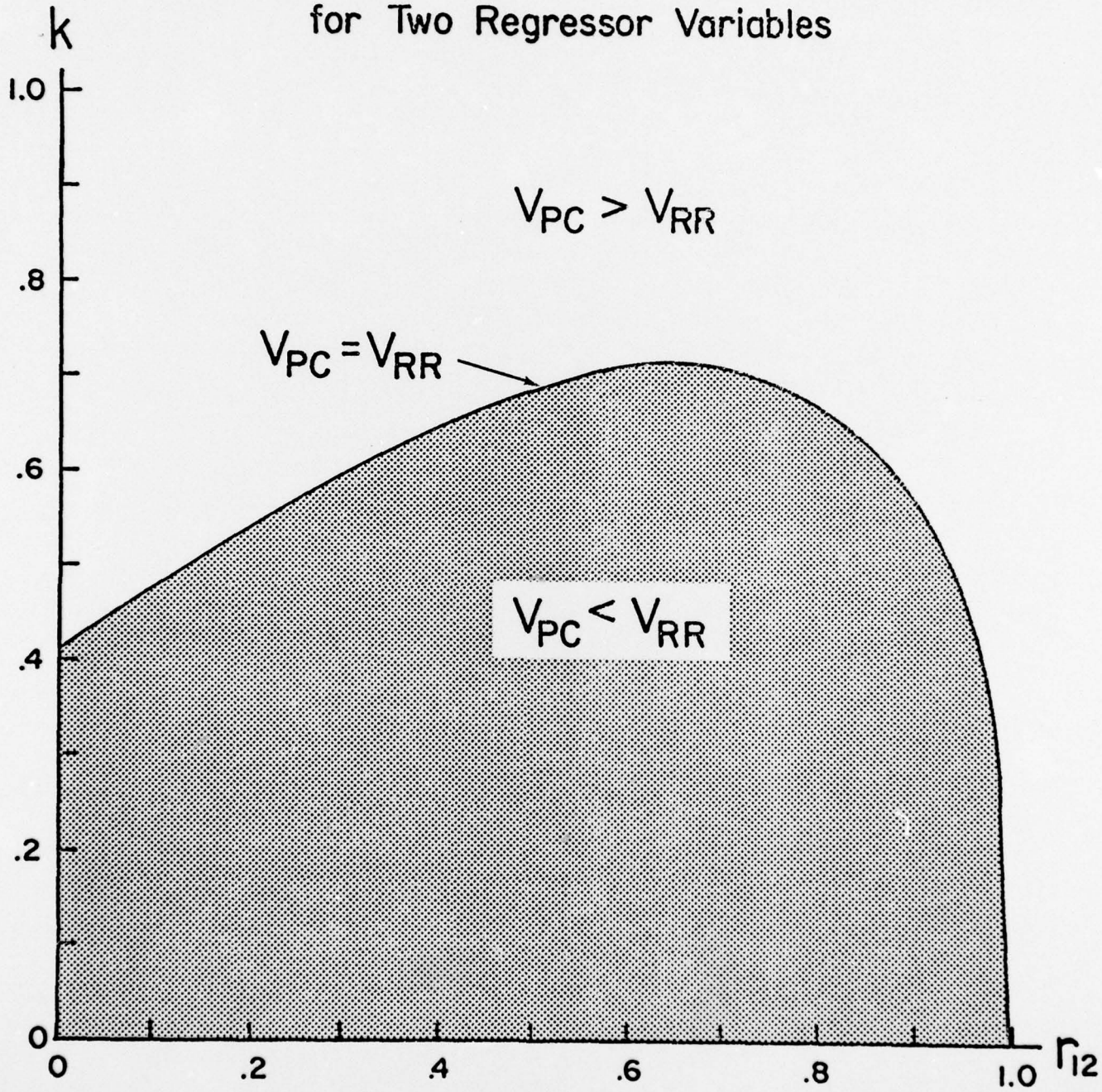


Figure 3.
Comparison of Integrated Bias With $p=2$, $r_{12}=0.95$.
(Shaded Region Indicates $B_{PC} < B_{RR}$;
Non-Shaded Region Indicates $B_{PC} > B_{RR}$).

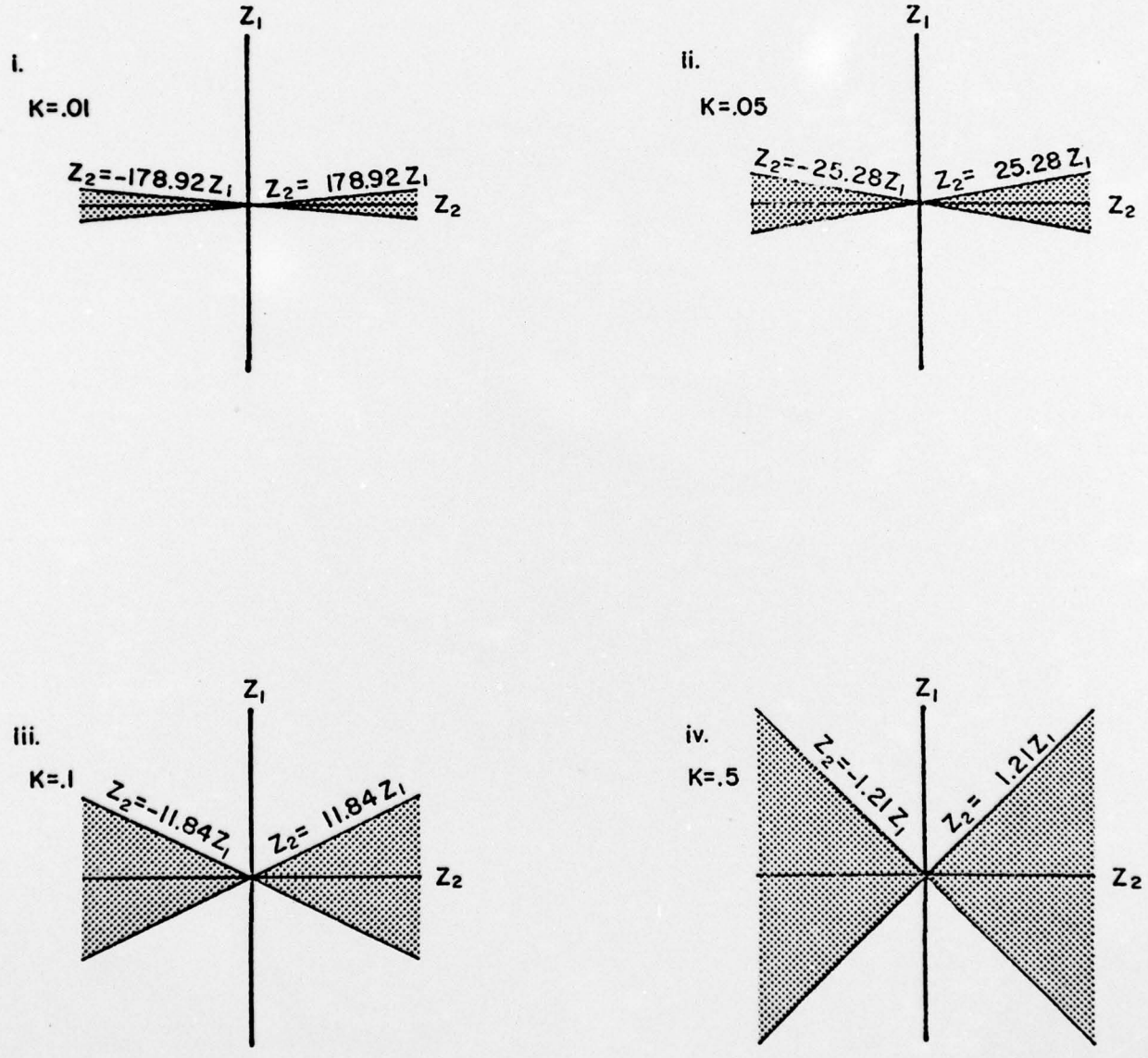


Figure 4.
Integrated Mean Square Errors for a Model
With $p=2$, $r_{12}=0.95$, $Z_2=0$.

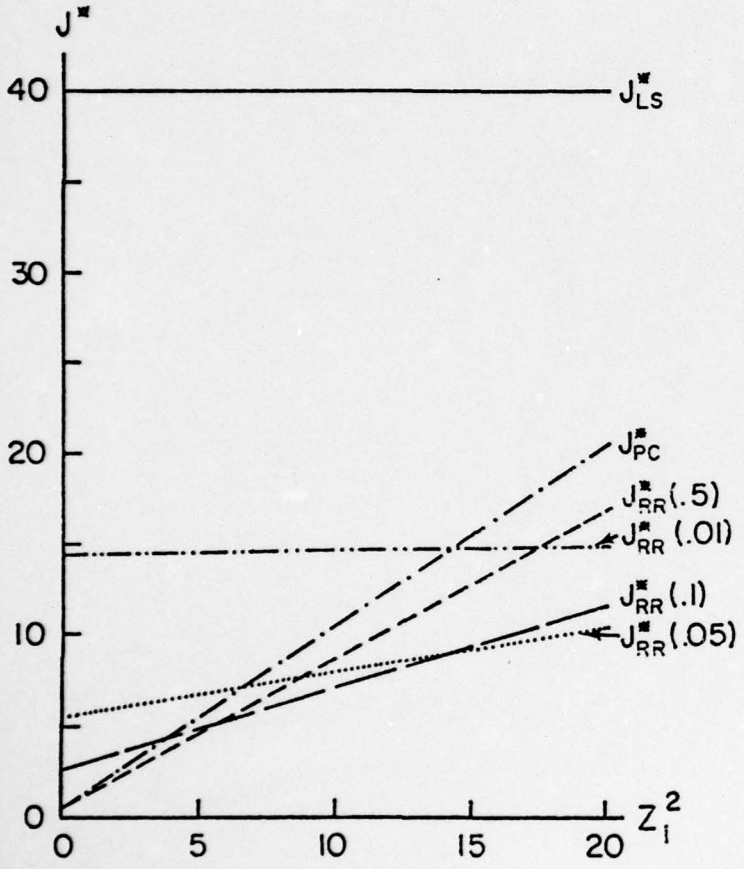
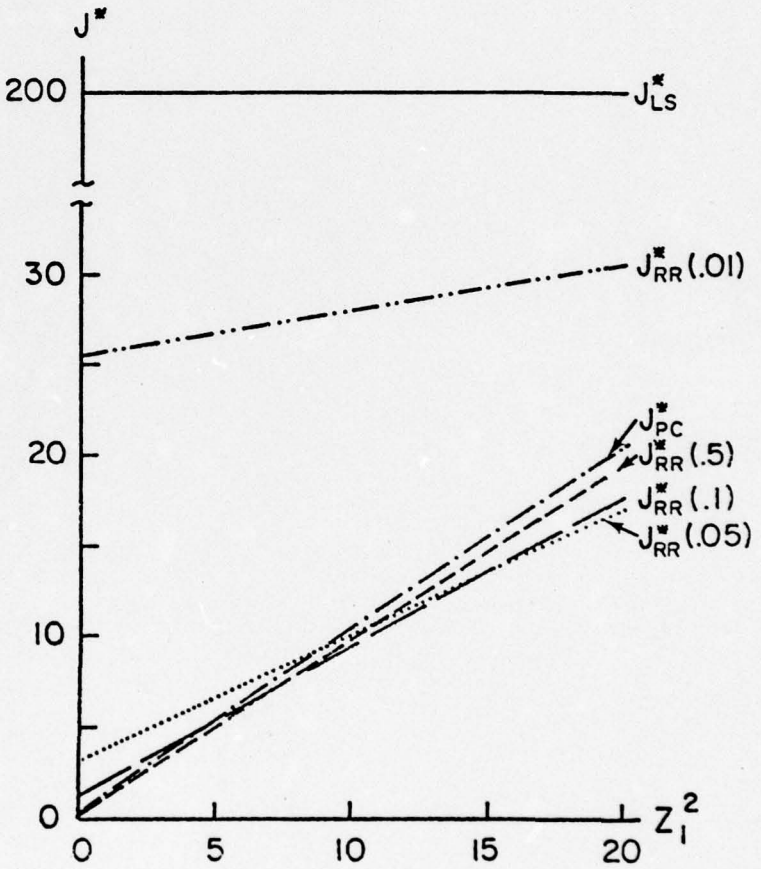


Figure 5.
Integrated Mean Square Errors for a Model
With $p=2$, $r_{12}=0.99$, $Z_2=0$.



REFERENCES

[1] Box, G.E.P. and Draper, N.R. (1959). "A Basis for the Selection of a Response Surface Design," J. Amer. Statist. Assn., 54, 622-54.

[2] Box, G.E.P. and Draper, N.R. (1963). "The Choice of a Second Order Rotatable Design," Biometrika, 50, 335-52.

[3] Cote, R., Manson, A.R., and Hader, R.J. (1973). "Minimum Bias Approximation of a General Regression Model," J. Amer. Statist. Assn., 68, 633-8.

[4] Davies, P. (1969). "The Choice of Variables in the Design of Experiments for Linear Regression," Biometrika, 56, 55-63.

[5] Draper, N.R. and Lawrence, W. (1965a). "Designs which Minimize Model Inadequacies: Cuboidal Regions of Interest," Biometrika, 52, 111-8.

[6] Draper, N.R. and Lawrence, W. (1965b). "Mixture Designs for Three Factors," J. Royal Statist. Soc. B., 27, 450-65.

[7] Draper, N.R. and Lawrence, W. (1965c). "Mixture Designs for Four Factors," J. Royal Statist. Soc. B., 27, 473-8.

[8] Gunst, R.F. and Mason, R.L. (1977). "Biased Estimation in Regression: An Evaluation Using Mean Squared Error," J. Amer. Statist. Assn., 72 (to appear).

[9] Helms, R.W. (1971). "The Predictor's Average Estimated Variance Criterion for the Selection-of-Variables Problem in General Linear Models," Institute of Statistics Mimeo Series No. 777, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

[10] Helms, R.W. (1974). "The Average Estimated Variance Criterion for the Selection-of-Variables Problem in General Linear Models," Technometrics, 16, 261-74.

[11] Hocking, R.R. (1976). "The Analysis and Selection of Variables in Linear Regression," Biometrics, 32, 1-50.

[12] Hoerl, A.E. and Kennard, R.W. (1970). "Ridge Regression: Biased Estimation for Non-orthogonal Problems," Technometrics, 12, 55-67.

[13] Karson, M.J. (1970). "Design Criterion for Minimum Bias Estimation of Response Surfaces," J. Amer. Statist. Assn., 65, 1565-72.

[14] Karson, M.J., Manson, A.R., and Hader, R.J. (1969). "Minimum Bias Estimation and Experimental Design for Response Surfaces," Technometrics, 11, 461-75.

[15] Kupper, L.L. and Meydrech, E.F. (1973). "A New Approach to Mean Squared Error Estimation of Response Surfaces," Biometrika, 60, 573-9.

[16] Kupper, L.L. and Meydrech, E.F. (1974). "Experimental Design Considerations Based on a New Approach to Mean Squared Error Estimation of Response Surfaces," J. Amer. Statist. Assn., 69, 461-3.

[17] Mansfield, E.R. (1975). "Principal Component Approach to Handling Multicollinearity in Regression Analysis," Ph.D. Dissertation, Department of Statistics, Southern Methodist University, Dallas, Texas.

[18] Marquardt, D.W. (1970). "Generalized Inverses, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation," Technometrics, 12, 591-612.

[19] Marquardt, D.W. and Snee, R.D. (1975). "Ridge Regression in Practice," Amer. Statist., 29, 3-19.

[20] Mason, R.L., Gunst, R.F., and Webster, J.T. "Sources of Multicollinearity in Regression Analysis," Comm. in Statist., 4, 277-92.

[21] Meydrech, E.F. and Kupper, L.L. (1976). "On Quadratic Approximation of a Cubic Response," Comm. in Statist., A5, 1205-13.

[22] Park, S.H. (1977). "Selection of Polynomial Terms for Response Surface Experiments," Biometrics, 33, 225-30.

[23] Webster, J.T., Gunst, R.F., and Mason, R.L. (1973). "Recent Developments in Stepwise Regression Procedures," Proc. of University of Kentucky Conference on Regression with a Large Number of Predictor Variables, W.O. Thompson and F.B. Cady (eds.), Department of Statistics, University of Kentucky, Lexington, Kentucky.

[24] Webster, J.T., Gunst, R.F., and Mason, R.L. (1974). "Latent Root Regression Analysis," Technometrics, 16, 513-22.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (19) (18) AFOSR - TR-77-0688	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* (6) SOME CONSIDERATIONS IN THE EVALUATION OF ALTERNATE PREDICTION EQUATIONS,		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Richard F. Gunst [redacted] Robert L. Mason		8. CONTRACT OR GRANT NUMBER(s) (15) AF-AFOSR 75-2871-75
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Department of Statistics Dallas, TX 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) 61102F 2304/A5 (17) A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE (11) 1977
		13. NUMBER OF PAGES ~~34~~ (12) 36p.
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Integrated mean squared error
prediciton equations
multicollinearity
biased estimation

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Prediction equations constructed from multiple linear regression analyses are often intended for use in predicting response values throughout a region of the space of the predictor variables. Criteria for evaluating prediction equations, however, have generally concentrated attention on mean squared error properties of the estimated regression coefficients or on mean squared error properties of the predictor at the design points. If adequate predication throughout a region of the space of predictor variables is the goal, →next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

20. neither of these criteria may be satisfactory in assessing the predictor. In this paper integrated mean squared error is used as a criterion to determine when the least squares, principal component, and ridge regression estimators of regression coefficients can produce satisfactory prediction equations in the presence of a multicollinear design matrix.